

5-1-2006

Using the Dynamic Indicators of Basic Early Literacy Skills to Identify Students At-Risk for Reading Difficulties

Rachael Sloan

Western Kentucky University

Follow this and additional works at: <http://digitalcommons.wku.edu/theses>



Part of the [Education Commons](#), and the [Psychology Commons](#)

Recommended Citation

Sloan, Rachael, "Using the Dynamic Indicators of Basic Early Literacy Skills to Identify Students At-Risk for Reading Difficulties" (2006). *Masters Theses & Specialist Projects*. Paper 481.
<http://digitalcommons.wku.edu/theses/481>

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.

USING THE DYNAMIC INDICATORS OF BASIC EARLY LITERACY SKILLS TO
IDENTIFY STUDENTS AT-RISK FOR READING DIFFICULTIES

A Specialist Project
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Specialist in Education

By
Rachael Noel Sloan

May 2006

**USING THE DYNAMIC INDICATORS OF BASIC EARLY LITERACY SKILLS
TO IDENTIFY STUDENTS AT-RISK FOR READING DIFFICULTIES**

Date Recommended May 2, 2005

Carl L. Myers

Director of Thesis

Melissa A. Hakman

Sam Beery

Elmer Gray 11-23-05
Dean, Graduate Studies and Research Date

Acknowledgements

I would like to thank my family and friends for their love and support through my work on this thesis as well as my work as a student. I would especially like to thank my mother for her never ending love and support in everything and anything I have attempted in my life. I would also like to thank Matt, who was an unlimited source of encouragement and ensured that I never took myself too seriously. I love you all and appreciate all of your support and love throughout the years.

I would also like to thank my committee members for all of their assistance, patience, and guidance with the completion of this project. I would especially like to thank Dr. Myers for his guidance, advice, and friendship throughout this project, as well as throughout my graduate school career.

Table of Contents

List of Tables.....	v
List of Figures.....	vi
Abstract.....	vii
Literature Review.....	1
Method.....	34
Results.....	38
Discussion.....	47
References.....	54

List of Tables

Table 1: Predictive Accuracy of Published and Adjusted DIBELS Cut-off Scores.....	44
Table 2: Comparison of DIBELS Total With and Without Minority Students.....	46

List of Figures

Figure 1: Classification of benchmark zones.....	30
Figure 2: Classification of students based on the published LNF benchmark of 15 correct letters named in one minute.....	39
Figure 3: Classification of students based on the published PSF cut-off of 7 correct phoneme segmentations in one minute.....	39
Figure 4: Classification of students based on the adjusted LNF cut-off at 31 correct letters named in one minute.....	41
Figure 5: Classification of students based on adjusted PSF cut-off at 11 correct phoneme segmentations in one minute.....	41
Figure 6: Classification of students using a cut-off of 17 correct initial sounds based on less than 5% false negatives.....	42
Figure 7: Classification of students based on DIBELS Total cut-off of 55.....	43

USING THE DYNAMIC INDICATORS OF BASIC EARLY LITERACY SKILLS TO IDENTIFY STUDENTS AT-RISK FOR READING DIFFICULTIES

Rachael Sloan

May 2006

60 Pages

Directed by: Carl L. Myers, Ph.D, Melissa Hakman, Ph.D, and Pamela Petty, Ed.D

Department of Psychology

Western Kentucky University

Reading is an essential skill that facilitates the development of later academic skills. Research has shown that phonological awareness at a young age is a strong predictor of later reading skills. As such, it is important to identify students who struggle with reading and phonological awareness at a young age. It is equally important that the tools used to identify early reading difficulties are appropriate and accurate for a given school district. The purpose of the current project was to examine the published benchmarks for the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in the identification of students as at-risk for later reading difficulties. In the project, 104 kindergarten students' DIBELS scores were compared to their California Achievement Test-Fifth Edition reading scores at second grade. Published benchmarks were then adjusted to decrease the number of students who were seen as on-target initially, but struggled with reading at a later age (false negatives). While the published DIBELS benchmarks were adequate as identifying most students as at-risk, the adjusted cut-offs identified fewer false negatives. This information is important for school psychologists and other educators in the process of identifying students who may be at-risk for later reading difficulties within their school district.

Literature Review

Reading is an essential component throughout many areas of a child's academic development and success. Ultimately, reading helps children acquire the knowledge and skills necessary to comprehend a variety of academic material at a level that is consistent with their ability (Torgensen, 2002). Reading acquisition appears to facilitate the development of later skills, and a deficit in early reading skills may lead to an inability to acquire later skills (Fleming, 2000). Without the necessary skill of reading, children are often left behind as their peers advance and achieve academic success. The students who are left behind often find themselves being labeled and referred for special education. According to the Digest of Education Statistics (2002), the percentage of students with learning disabilities comprises nearly half (45.2%) of those receiving special education services. The importance of reading and its vital components are becoming more apparent as it enters into public policy and debate as evidenced by President Bush's "No Child Left Behind" initiative, which states that all children should read at or above grade level by grade three (Department of Education, 2004).

Importance of Early Reading Intervention

Without the appropriate intervention, children who are poor readers or who have difficulty reading in the first grade have a .88 probability of remaining a poor reader at grade four (Juel, 1988). Therefore, it is essential that those children labeled as "poor readers" are identified early and a proper intervention plan is employed. Without proactive measures, these children are very likely to continue to struggle academically and fall farther and farther

behind their peers, who succeed at reading, and who continue to advance and achieve—a concept known as the Matthew Effect (Stanovich, 1986). The Matthew Effect refers to the Gospel of Matthew which states that the rich get richer as the poor get poorer. Because reading is such a basic and necessary skill, children who lack the skill are unable to obtain, or obtain at a slower rate, all other academic skills. In contrast, those who possess the basic reading skills are able to grasp all other academic skills at a faster rate, thus widening the gap between the two cohorts. In addition, the Matthew Effect can be used to explain other findings, such as why academic progress for some children can be extremely slow, why teachers may not see a significant change in a child's general classroom success, and why simply teaching phonemic awareness skills to older children may not have the desired impact (Hempenstall, 1997). Furthermore, children who encounter academic difficulties are more likely to also experience behavioral and social problems, as well as a loss of self-esteem (Coyne-Martinson, 2000). The implications of these findings further the argument for early reading instruction and intervention.

Good, Simmons and Kame'enui (2001) noted that in order for a reading intervention to be successful in a prevention-oriented system it must adhere to the following standards: (a) intervene early and strategically during critical windows of reading development, (b) develop and promote a comprehensive system of research-based instruction of core curriculum, (c) use and rely on formative, dynamic indicators of student performance to identify needs, allocate resources, and design and modify instruction, and (d) address reading failure and reading success from a schoolwide, systematic perspective. With these elements in place, an intervention program is likely to prove more successful and beneficial in early prevention of

reading difficulties. Such standards are illustrated in intervention studies such as those conducted by Torgensen, Rashotte, and Alexander (2001) and Torgensen, et al. (1999).

The intervention study performed by Torgensen et al. (2001) provided two different types of intervention instruction to a group of sixty 8- to 10-year-olds who were receiving special education instruction in reading prior to the start of the study. Both instructional methods consisted of instruction in phonemic decoding and orthographic reading skills; the only variation was with the method of instruction as well as the amount of time spent on various instructional activities. Both groups received instruction divided into two, 50-minute sessions a day; however, one group received instruction in phonemic awareness and phonemic decoding, while the second group received the same instruction with more emphasis on the application in meaningful text. During the 16 months that the children had received special education instruction prior to the study, they had made only slight improvements in their standard scores on the Woodcock Reading Mastery Test-Revised. However, with the instructional methods implemented by the authors, the children made dramatic improvements. Both methods of instruction proved successful in increasing students' reading ability, and the gain remained stable at the time of a two-year follow up. The authors concluded that both methods were equally successful as no differences existed between the groups at the time of the follow up (Torgensen et al., 2001). While the differences in instruction did not affect the gains made by the students, the interventions proved highly successful, in terms of reading gains, when compared to the students' previous gains in the 16-month period of special education instruction. The findings illustrate that an intensive and focused approach is more successful than general special education services.

Another study conducted by Torgensen et al. (1999), examined three types of instructional methods, two of which varied in intensity while the third supported regular classroom instruction. The study consisted of 180 kindergarten students who had scored the lowest combined scores (bottom 12%) on a letter naming task and phonemic awareness task and also had a Verbal IQ score of above 75. The students were assigned to one of four groups: (a) control group, (b) support for regular classroom, (c) embedded phonics, or (d) phonological awareness plus synthetic phonics. The embedded phonics group consisted of grapheme-phoneme instruction taught within the context of the activities rather than explicitly. The phonological plus synthetic phonics group received explicit instruction in phonemic awareness in addition to skills for determining the phonemes in words. The study lasted over a two and a half year period into the second grade. The findings of the study indicated that the group that received the most explicit form of phonemic instruction at the kindergarten level proved more successful at reading in second grade than the embedded phonics group. Both the phonemic explicit group and embedded phonics group showed greater gains in reading ability than the group that received support in the regular classroom or the group that received no intervention at all (Torgensen et al., 1999).

While early intervention may be a key variable to the development of a successful reader, the type of intervention is equally essential. The most common cause of difficulties in acquiring early word reading skills is weaknesses in the ability to process the phonological features of language (Liberman, Shankweiler, & Liberman, 1989). To grasp the importance of phonological intervention, it is essential to first understand the nature of phonological awareness.

Phonological Awareness

Phonological awareness refers to one's awareness of the structure of words in the written language (Torgensen, Wagner, & Rashotte, 1994). In order for a child to be successful with phonological skills, he or she must be able to attend to the individual letters of the words, their sequencing, and their phonological translations (Adams, 1990). Adams noted that children's knowledge and level of phonological awareness grows in a hierarchical manner. Adams presented the first level of the hierarchy as the knowledge of letter-sound correspondences. This level is simply the child's knowledge that certain letters correspond with certain sounds. Next children learn that letters placed in a particular sequence make up parts of a word, which are then connected to form a whole word—the final level of the hierarchy. However, Adams goes on to note that the process of phonological awareness is not as upwardly directional as it may first appear. A student must possess the simplest of knowledge (what a letter is) before he or she can associate it with higher meaning. In addition, students must not only possess basic understanding of what a phoneme is but also the many conditions under which various phonemes occur. Students must be able to internalize the rules and variations which accompany phonics to be successful readers (Adams, 1990). In sum, while Adams believes that phonological awareness is essential for a student to be a successful reader, it does not necessarily occur in a sequential manner.

Mercer and Mercer (2001) also presented levels of hierarchical phonological awareness skills. While Adams' (1990) hierarchy was more general, Mercer and Mercer (2001) believe the skills progress in a more structured hierarchical manner. Mercer and Mercer's hierarchy is as follows:

Word. Word level knowledge occurs at a young age. Children are able to produce in isolation words which had previously been heard in conjunction with other words. The ability to tap out or count the number of words heard in a sentence demonstrates word level knowledge (Mercer & Mercer, 2001).

Syllable. The ability to separate words into syllables is a fairly simple task and can be measured through tapping out the number of syllables in a word (Mercer & Mercer, 2001).

Onset and Rime. Onset and rime refers to one's ability to separate syllables into smaller units. Onset refers to the part of the syllable that precedes a vowel. The rime is the rest of the word following the onset (Mercer & Mercer, 2001).

Phoneme. Phonemic awareness is at the top of the phonological hierarchy. Phonemic awareness applies to the sound associated with a given phoneme, while phonological awareness refers to graphemes or print letters of the word (Mercer & Mercer, 2001). However, the two terms are often incorrectly used interchangeably. Phonemes are the smallest unit of sound in a word. In its most basic terms, phonemic awareness is the awareness that words are made up of individual units of sounds (Good and Kaminski, 1996; Hempenstall, 1997; Snider, 1995; Snider, 1997; Stahl & Murray, 1994; Yopp, 1992). A more extensive definition is offered by Yopp (1992), who defines phonemes as abstract units of speech that do not have physically definable boundaries and are influenced by their phonological context. Phonemic awareness is simply the awareness of these properties. It refers only to the spoken language, and it is a separate entity from the awareness of *meanings* of words within a language (Good and Kaminski, 1996; Hempenstall, 1997; Stahl & Murray, 1994). In sum, phonemic awareness is the conscious awareness that words are made up of smaller units of sounds, and that those units of sound can be manipulated and altered to form

other words. Such awareness is not necessary or fundamental to speak or understand the spoken language; however, it is necessary in the understanding of the written word and therefore in the ability to read (Good and Kaminski, 1996; Hempenstall, 1987; Snider, 1997; Stahl & Murray, 1994).

The ability to read and be phonemically aware goes hand in hand. For a child to be successful in acquiring basic reading skills, the child must possess phonemic awareness. In turn, a child who is successful in reading then garners a better understanding of language and words (Yopp, 1992). Phonemic awareness is not only essential to reading but also spelling, because the English language is made up of letters which correspond to sounds that comprise phonemes (Snider, 1995). Each knowledge component builds on the last until the skill of decoding has been acquired.

There are many implications associated with a child's performance on phonemic awareness tasks. Performance on the phonemic awareness tasks of blending and segmenting is highly correlated with later acquisition of early reading skills (Wagner & Torgensen, 1987). A student's performance on phonemic awareness tasks provides an abundance of information on a student's skill level as well as providing insight into later reading achievement; however, it is not the task itself that is important to reading skills. As noted by Adams (1990), it is neither the ability to distinguish phonemes nor the ability to distinguish between phonemes that are essential, but the awareness that phonemes exist as malleable segments of speech.

In order for children to acquire basic literacy skills, they must possess phonemic awareness. Therefore, a child who does not possess, or possesses limited, phonemic awareness skills will likely be a poor reader; whereas, a child who possesses basic phonemic

awareness skills will likely be a successful reader (Fleming, 2000; Hempenstall, 1997; Yopp, 1992). Because phonemic awareness plays such an essential role in reading, it is important to determine whether training in such tasks would improve a child's chance at becoming a successful reader. Griffith and Olson (1992) cited research that indicated that training in phonemic awareness had a positive effect on reading success; the training was even more successful when paired with training in the alphabetic principle. In order for the instruction and training to be successful, it must be explicit and presented in a logical sequence (Griffith, 1992). The success of phonemic awareness training programs on reading provides even more evidence to support the hypothesis that phonemic awareness is a precursor to reading success. According to Adams (1990):

faced with an alphabetic script, the child's level of phonemic awareness on entering school may be the single most powerful determinant of the success she or he will experience in learning to read and of the likelihood that she or he will fail. (p. 304)

Therefore, phonemic awareness may be the single best predictor of reading and spelling success (Adams, 1990; Lundberg, Olofsson, & Wall 1980; Mann, 1993; Stanovich, 1986; Tunmer & Nesdale, 1985).

Hierarchy of Phonemic Awareness Tasks

Phonemic awareness is not a singular ability; it is a general construct that can be demonstrated in different ways and measured by different tasks (Snider, 1997). In addition, skill or success in one area does not necessarily result in success in all areas of phonemic awareness (O'Connor, Jenkins, Leicester, & Slocum, 1993). Phonemic awareness is evident in many rudimentary tasks such as rhyming and alliteration which are often common in preschool and early school years (Snider, 1997). Thus, phonemic awareness appears to be an

inherent skill, at least to some degree, as children are not instructed in reading at the preschool level, nor are they taught the rules of rhyming. Yet children as young as preschoolers are able to detect rhymes when recited orally. Phonemic awareness is also apparent in more sophisticated tasks such as determining beginning and ending word sounds as well as segmenting words into their individual phonemes (Adams, 1990). By understanding how children develop phonemic awareness skills, interventions can be tailored for a student's skill level as well as provide a guideline for student progress. Adams (1990) described a set of tasks which increases in difficulty and requires higher and higher levels of phonemic ability.

Knowledge of nursery rhymes. Knowledge of nursery rhymes is simply the ability to recognize by ear the similarity of the sounds of certain words. Early knowledge of nursery rhymes has been shown to be related to later phonological skills as well as later reading acquisition (Adams, 1990). Knowledge of nursery rhymes is the most simplistic of the phonemic awareness tasks and has been recognized even in very young children.

Oddity tasks. In oddity tasks, children are orally presented with a set of three or four words and asked which word does not belong or is different based on either the beginning, middle, or ending sounds of the words. As noted by Adams (1990) the task is simple in that it only requires the child to rely on his or her ability to detect like and unlike sounds; therefore, it is especially useful with preschool aged children.

Blending tasks. In blending tasks, the examiner orally presents the child with segments of a given word such as /m/ . . . /a/ . . . /p/. The child is then asked to blend the sounds together so that it makes a word (Adams, 1990). Blending tasks are still considered an easier measure than other more involved and complex tasks.

Syllable-splitting tasks. The syllable splitting tasks are more difficult than the previously mentioned methods. The children are required to separate the first syllable of the word from the remainder of the word and may also be asked to pronounce the separated phoneme. The task requires that the child listen to the sounds of the word and be aware that the initial sound can be separated from the rest of the word and exist on its own (Adams, 1990).

Phoneme manipulation tasks. The phoneme manipulation task is similar to the syllable-splitting tasks. In both tasks a segment of the word is removed, typically the first. However, the manipulation task requires the child to pronounce the word that *remains* after the phoneme of choice has been removed. For example, “Say *hill* without the /h/.” In order for the child to be successful in picking out relevant phonemes from a word, the child must have some knowledge of phonemes. More advanced skills are required to actually manipulate the word, as in deleting and reordering the phonemes (Adams, 1990).

Phoneme segmentation tasks. The most difficult of the tasks which measure phonemic awareness, phoneme segmentation tasks, require the child to break a word down into the contributing phonemes. Within this category are different methods of measuring a child’s ability to segment phonemes. Perhaps the most prevalent method is the tapping task in which the child is presented with a word and then asked to tap out the number of phonemes in each syllable. For instance, given the word “*map*” the child should tap three times for the phonemes /m/.../a/.../p/ (Adams, 1990). Phonemic segmentation fluency also provides a method of assessing this skill. The child is presented with a word orally and then asked to repeat the word in its segmented syllables (Good & Kaminski, 2002). These tasks are more complicated than others because there is no way to know that a word is composed

of separate phonemes unless it has somehow been taught. Phoneme segmentation tasks require that the child not only know that a syllable can be segmented into phonemes but also be aware of what a phoneme is and how it is used in the formation of words (Adams, 1990).

Studies of the Association Between Reading and Phonemic Awareness

Numerous research studies have been conducted in order to further understand the predictive nature of both phonemic awareness and phonological awareness tasks. Studies found that phonemic and phonological awareness are strong predictors of future reading ability. In addition, most of the studies found that the tasks were successful at predicting reading ability in the early school years. The following is a review of relevant findings.

Lundberg et al. (1980) conducted a study to further examine the relationship between metalinguistic skills, general abilities, reading skills, and spelling skills. The study included word analysis, as measured by multiple phonemic segmentation tasks, as well as knowledge of phoneme positions. The study initially tested 200 kindergartners. Data were available from 143 students at the time of follow-up in first grade, and data from 133 students were available in the first semester of second grade. The results indicated that the most powerful predictors of reading and writing skills were the phoneme segmentation tasks and the ability to reverse phoneme order—phoneme manipulation.

In another study of phonological awareness (which included tasks measuring phonemic awareness), 525 kindergartners with a mean age of 5 years and 3 months were tested during kindergarten and 479 at the end of first grade (Share, Jorm, Maclean, & Matthews, 1984). Reading achievement was assessed through the use of sight words, scrambled story words, nonsense words, the Neale Analysis of Reading Ability, and spelling words. The students were tested on numerous individual ability tasks including a phoneme

segmentation task. Of the tasks presented, phoneme segmentation was the strongest predictor of first grade reading achievement.

A study by Tunmer and Nesdale (1985) set out to further illustrate the connection between phonemic segmentation and learning to read by presenting first graders with tests of phonemic segmentation (tapping), verbal ability, and reading achievement. Sixty-three first graders participated in the study with the average age being 6 years, 2 months at the start of testing. Reading ability was assessed through three subtests of the Interactive Reading Assessment System, which included real word decoding, pseudoword decoding, and reading comprehension. The authors concluded that the phonemic awareness tasks of tapping out syllables was strongly correlated with decoding ability, and decoding ability was strongly correlated to reading comprehension. To further examine the relationship between the skills, individual patterns were examined. All of the students who did well on the decoding tasks also did well on the tapping test. All of those who failed the tapping test also failed the decoding task. Therefore, it appears that phonemic awareness is a necessary skill prior to being able to decode. There was also a group of children who were able to segment or tap out words but were unable decode; therefore, the children must have been lacking an additional skill.

The predictive nature of phonemic awareness tasks has also resulted in reliability and validity studies of measures designed to assess the skills needed to become a successful reader. One such study, performed by Yopp (1988), used a cohort of 98 kindergartners with an average age of 5 years, 10 months who were tested on 10 tasks of phonemic awareness, as well as an invented spelling test, which required the students to decode novel words. The 10 tasks included an auditory discrimination test, a phoneme blending test, a phoneme counting

test, two phoneme deletion tests, a rhyming test, two phoneme segmentation tests, a sound isolation test, and a word-to-word matching test. A learning test was also included in the study to determine the predictive validity of each measure. An examination of the tasks in relation to the learning criterion, which was the ability to read novel words, revealed that each task had moderate to high correlations. The Sound isolation task had the highest correlation followed by the two phoneme segmentation tests. The study illustrated that a variety of phonemic awareness tasks have predictive validity in predicting future reading ability.

In another study, a group of 100 children were pooled from two public schools with the mean age of 71.1 months with ages ranging from 64 months to 81 months (Mann, 1993). The children were initially tested in May of their kindergarten year on two phonemic awareness tasks—phoneme segmentation and the invented spelling test which infers phonemic awareness from kindergartners early attempts to “write” spoken words. The author found that children who can determine which of four words starts with a different sound were more likely to become better readers than those who were unable recognize the dissimilar sounds. This study once again demonstrates how knowledge of phonemes and phoneme segmentation skills can be used to predict later reading ability.

In another longitudinal study, MacDonald and Cornwall (1995) followed-up with 24 teenagers with a mean age of 17, who had previously participated in a study 11 years earlier that assessed their phonological analysis and reading and spelling abilities. The original study was conducted in 1982 with a group of 58 kindergartners who were given the Auditory Analysis Test (AAT), the Picture Vocabulary Test, and the Reading and Spelling subtests of the Wide Range Achievement Test. During follow-up, the 24 students were again presented

with the same three measures along with the Word Attack and Passage Comprehension subtests of the Woodcock Reading Mastery Test-Revised. The authors found that phonological awareness, as measured by the AAT, was both a concurrent and long-term predictor of word identification and spelling skills as assessed at both points of the study (McDonald & Cornwall, 1995). In addition, the phonological skills of the subjects at age 6 were indicative of word identification and spelling skills at age 17. Perhaps most importantly, phonological tasks were also shown to be more predictive of reading and spelling skills than either socioeconomic status or vocabulary knowledge.

An examination of the Yopp-Singer Test of Phoneme Segmentation was conducted by Yopp (1995). The test is designed to measure a child's ability to separate words into its individual phoneme components and articulate the phonemes in the correct order back to the examiner. Scores on the task serve to evaluate a student's overall level of phonemic awareness. A student who is able to segment all or almost all of the words is considered phonemically aware, a student who is able to segment some or around half the items is viewed as possessing emerging phonemic awareness, and a student who is able to segment only a few words or no words is considered phonemically unaware. The predictive validity of the instrument was assessed through the collection of reading achievement data that began with a group of kindergartners and concluded when they were in sixth grade. Performance on the Yopp-Singer task compared to follow-up measures of reading, spelling, comprehension, vocabulary, and word attack tasks, indicated a moderate to strong relationship indicating significant predictive validity (Yopp, 1995).

A longitudinal study conducted by Snider (1997) examined the relationship between phonemic awareness and reading achievement. Seventy-three kindergartners' scores on a test

of phonemic awareness were compared with their scores on a standardized reading achievement test at the end of second grade. Five tasks were employed to measure phonemic awareness: rhyme supply, sound oddity tasks, blending tasks, phoneme segmentation, and phoneme manipulation. Snider found that the rhyme supply and sound oddity tasks were not very predictive of later reading ability while the other three tasks predicted reading achievement equally well. The authors found that mean scores on the test of phonemic awareness were low and standard deviations were high, indicating a high level of variability in performance. The authors noted that a student's poor performance on phonemic segmentation and manipulation tasks was not necessarily indicative of the student's later reading performance due to the high variability. Many students who scored low on the phonemic segmentation and manipulation tasks in kindergarten went on to become successful readers. It is unclear as to how much regular classroom instruction in reading may account for the students' later success as readers. However, low scores at the end of kindergarten may serve as an indicator for reading failure in first grade (Snider, 1997).

A study conducted by Chafouleas and Martens (2002) examined the efficiency and sensitivity of five accuracy-based phonological awareness tasks including rhyme-providing, sound-providing, blending, segmentation, and deletion. Sixty-seven students in kindergarten and first grade were tested on the five tasks with each task containing 10 items. The results of the study indicated that the five phonemic awareness tasks were capable of showing student growth as well as predicting future reading ability. All five measures were more predictive of later reading ability when used in kindergarten than in grade one. The most sensitive measures were the segmentation, rhyme-and sound-providing tasks with segmentation appearing to be the most sensitive measure.

In sum, tasks that measure phonemic awareness appear to be strong predictors of later reading skills. In addition, Tunmer and Nesdale (1985) noted that phonemic awareness was a necessary skill prior to the ability to understand phonics—which is an important ability in the reading process. Phonemic awareness tasks have also been shown to be more predictive of later reading skills than socioeconomic status or even vocabulary knowledge (MacDonald & Cromwell, 1985). MacDonald and Cromwell (1985) also found in their longitudinal study that scores on tasks of phonemic ability were indicative of reading skills 17 years later. While the studies reviewed indicate phonemic awareness tasks are strong indicators of later reading ability, variability of scores—as in the study conducted by Snider (1997)—may over-identify students as at-risk in the development of reading skills. More specifically, a young student who scores low on a phonemic awareness task may go on to become a successful reader.

Dynamic Indicators of Basic Early Literacy Skills

A popular measure which incorporates an assessment of phonemic awareness is called the Dynamic Indicators of Basic Early Literacy Skills ([DIBELS] Good & Kaminski, 2002). The measures were originally developed as a downward extension of Curriculum-Based Measurement (CBM). CBM is a set of standard, short-duration, fluency measures of reading, spelling, written expression, and mathematic computation developed to serve as indicators of student achievement (Shinn & Bamonto, 1998). Good and Kaminski (1996) developed the DIBELS measures at the University of Oregon. The original DIBELS measures consisted of 10 brief tasks designed to measure a student's progress over time as well as identify those children who are in need of reading intervention (Elliott, Lee, & Tollefson, 2001). The DIBELS measures three of the five areas of early literacy described in *Big Ideas in Beginning Reading* (National Reading Panel, 2000). The five essential areas of

reading were phonemic awareness, the alphabetic principle, fluency with connected text, vocabulary, and comprehension. Currently, the sixth edition of the DIBELS consists of six standardized, individually administered measures designed to assess the first three of the five listed areas (phonemic awareness, alphabetic principle, and fluency with connected text). The number of tasks given is dependent on the child's grade in school and the point in the school year in which the child is being tested (fall, winter, or spring). The DIBELS tasks are as follows:

Initial Sound Fluency. Initial Sound Fluency (ISF), which was formerly known as Onset Recognition Fluency on previous editions of the DIBELS, assesses a child's ability to differentiate the initial sound of a given word from the rest of the word (Kaminski & Good, 1998). The task is designed to measure phonemic awareness. The examiner presents the child with a page consisting of four pictures. The examiner then states the names of the pictures and asks the child to point to or state the picture that begins with a certain sound orally presented by the examiner. For example, the examiner would say, "This is sink, cat, gloves, and hat. Which picture begins with /s/?" This process is repeated for two more pictures on the same page. For the fourth and final picture the child is asked to produce the initial sound for an orally presented word that corresponds with the picture. Scores are calculated based on the number of correct responses provided in one minute.

The ISF task is presented at the preschool age as well as in the fall and winter of the kindergarten year. For the winter kindergarten administration the benchmark goal (an indication of where the child should be performing at that time of the year for that grade) is 25 to 35 correct onsets per minute (Good & Kaminski, 2002). According to the DIBELS manual (Good & Kaminski, 2002), the alternate form reliability for the Initial Sound Fluency

task is .72 in January of the kindergarten year. A predictive validity coefficient of ISF with reading ability in the spring of the first-grade year as measured by a curriculum-based measure was found to be .45. In addition, the predictive validity of the ISF task in kindergarten with the Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster Score was .36 (Good & Kaminski, 2002).

Phoneme Segmentation Fluency. The Phoneme Segmentation Fluency (PSF) task, which also measures phonemic awareness, is intended for administration in the winter and spring of kindergarten as well as the fall, winter, and spring of first grade. PSF assesses the student's ability to segment or break down three-and four-phoneme words into the individual phoneme components. To administer the PSF task, the examiner orally presents words one at a time, and then requires the child to break the word down into individual phonemes and recite the phonemes to the examiner (Good & Kaminski, 2000). For example, the examiner would say, "sat," and a correct response by the student would be "/s/ /a/ /t/." A student's score is determined by the number of correct segmentations provided in one minute (Good & Kaminski, 2002). The benchmark for the task is 35 to 45 correct phoneme segmentations in one minute for the spring of kindergarten and fall of first grade (Good & Kaminski, 2002.) The PSF task has been found to be a good predictor of later reading ability (Kaminski & Good, 1996). The two-week alternate-form reliability for the measure was estimated at .88 for the kindergarten year. In addition, the predictive validity coefficient was .68 with the Woodcock-Johnson Psycho-Educational Battery Total Reading Cluster score in first grade (Good & Kaminski, 2002).

Nonsense Word Fluency. Nonsense Word Fluency (NWF) is designed to measure the "big" idea of the alphabetic principle. Through the use of a list of nonsense words, the task

requires children to possess knowledge of letter sound correspondence as well as the ability to blend letters into words. The NWF task is appropriate for students in the winter of their kindergarten year to students in the fall of their second grade year. The benchmark for this task is 50 correct letter sounds per minute by first grade (Good & Kaminski, 2002).

Oral Reading Fluency. Oral Reading Fluency (ORF) measures fluency with connected text, and is appropriate for students in the winter of their second grade year to the spring of the third grade year. The task requires the child to read a standardized set of three passages. Students' scores are the number of correct words read per minute. The benchmark goals for ORF are 40 words per minute in the spring of first grade, 90 in the spring of second grade, and 110 in the spring of third grade.

Letter Naming Fluency. Letter Naming Fluency (LNF) is also included in the DIBELS measures but does not fall into one of the three areas of "big ideas" of literacy. Each student is presented with a page of upper and lower case letters arranged in a random order. The child is then asked to name as many letters as possible within one minute. The student's score is the number of correctly named letters. The task is designed for students in the fall of the kindergarten year to the fall of the first grade year. A benchmark for the LNF task is not provided in the DIBELS manual, although the authors suggest that students scoring in the lowest 20% are at risk for difficulty achieving early literacy benchmarks.

Word Use Fluency. In the Word Use Fluency (WUF) task, students are asked to use a given word in the context of a sentence. The task does not appear to assess any of the five areas of early literacy; more research is needed to determine the WUF task's link with the big ideas of literacy. Again, a benchmark is not provided; however, students scoring between the 20th and 40th percentiles should be considered at some risk for having difficulty achieving

early literacy benchmarks, while those scoring below the 20th percentile are considered at-risk. The task is seen as an appropriate measure for the fall of the preschool year through the spring of the third-grade year (Good & Kaminski, 2002).

DIBELS Reliability and Validity Studies

In order to accurately assess a student's early literacy skills in phonemic awareness, it is essential that the measures be reliable and valid. One study conducted by Kaminski and Good (1996) examined the reliability, validity, and sensitivity of three DIBELS measures: PSF, LNF, and Picture Naming Fluency (students are asked to name as many given pictures as they can with in one minute). The Picture Naming Fluency (PNF) is not currently included in the latest edition of the DIBELS. The study involved one cohort of 37 kindergartners and one cohort of 41 first graders for a total of 78 participants. The students were randomly divided into two groups, a monitored group in which the students were given the three DIBELS measures two times a week for nine weeks, and a nonmonitored group which were tested on the measures at the beginning and end of the nine-week period.

The criterion measure was a Curriculum-Based Measure (CBM) of reading, which is now also referred to as Oral Reading Fluency in the current edition of the DIBELS. Students were asked to read a selected passage aloud for one minute, and the total number of words read correctly in one minute was recorded. The data collection aspect of the study was divided into three phases. During the first phase, all children were given the three DIBELS measures along with a measure of general intellectual ability. The students' teachers also completed a teacher rating scale to obtain ratings of reading achievement and readiness. In addition, the kindergarten and first-grade students were tested on two age appropriate reading measures. During the second phase of the study, children in the monitored group were

administered the DIBELS measures two times a week over the nine-week period. Additionally, the first graders were tested twice a week using the CBM reading measurement. For the final phase of data collection, all measures were readministered except for the test of general intelligence (Kaminski & Good, 1996).

Kaminski and Good (1996) noted that the first grade cohort scored significantly higher on all three DIBELS measures than the kindergarten cohort, suggesting that children's performance on the measures changes dramatically from the end of the kindergarten year to the end of the first-grade year. This finding indicates the DIBELS measures were sensitive to changes and growth from kindergarten to first grade with the PSF task having the highest rate of growth. All three of the DIBELS measures had higher reliability and validity coefficients for the kindergarten cohort than for the first-grade cohort. In addition, the LNF task had the highest reliability for both cohorts. In general, the DIBELS measures appeared to be valid and reliable for assessing reading ability.

Elliot (1997) also conducted a reliability and validity study of the DIBELS seeking to expound on the previous studies of the measures. The study set out to examine the adequacy of four DIBELS measures: Letter Naming Fluency, Phoneme Segmentation Fluency, Initial Sound Fluency, and Sound Naming Fluency (SNF) which was viewed as an experimental DIBELS task at the time (Elliot, 1997). The study included 75 kindergartners from a small Midwestern town. Elliot (1997) found that the results of the experiment partially replicated the previous work of Kaminski and Good (1996); however, concurrent validity estimates for LNF and PSF were generally found to be weaker. In addition, in the original study of Kaminski and Good, the authors found that the reliability for LNF was the strongest for kindergartners and first graders and PSF appeared to be the most sensitive to change, while

the Elliot study found the LNF and experimental SNF tasks to be the strongest measures. Additionally, Elliot stated that the PSF task required more precise scoring and standardization procedures.

Another study conducted by Elliot and colleagues evaluated the reliability and validity of modified versions of the LNF, SNF, ISF, and PSF tasks (Elliot et al., 2001). The study used a sample of 75 kindergartners who were administered the measures in two-week intervals at the end of the kindergarten year. The authors' purpose was to assess the technical adequacy of the four measures, as well as correct for floor effects encountered in their earlier study. For the study, the ISF and PSF tasks were modified and renamed initial phoneme ability (IPA) and phoneme segmentation ability (PSA), respectively, to denote their modifications (DIBELS-M). The modification of tasks for the IPA and PSA measures included selecting easier words from a kindergarten reading text as well as providing students with additional response time. The authors reasoned that the modification of the measures would serve as better indicators of student phoneme accuracy than fluency rates.

In addition to completing the DIBELS-M measures, Elliot et al. (2001) had the students complete four other instruments inclusive of two clusters (Broad Reading and Skills) on the Woodcock-Johnson Psychoeducational Assessment Battery—Revised, the Test of Phonological Awareness—Kindergarten form, and the Kaufman Brief Intelligence Test. The DIBELS-M measures served as the predictor measures while the other tests were the criterion measures. In addition, the authors combined the DIBELS-M measures into subgroups based on correlation patterns between the measures. The composite scores were designated as “Fluency,” “Ability,” and “Total.”

The study examined three different reliability estimates: interrater, test-retest, and alternate form (Elliot et al., 2001). For interrater reliability, the scores generally fell in the mid to upper .80s for both individual and composite scores. Estimates for interrater reliability were the highest for LNF and the weakest for SNF. Test-retest reliability was garnered through correlations between the second and third testing sessions; estimates ranged from .74 (IPA) to .90 (LNF). Alternate form reliability ranged from .80 to .91, with the exception of IPA which had a coefficient of .64. The concurrent validity estimates for the DIBELS-M indicated that the instrument was able to account for no less than 16% of the variance and, in most cases, between 30 and 40% of the variance in scores on achievement measures (Elliot et al., 2001).

The authors concluded that their findings were consistent with those found by Kaminski and Good (1996). The LNF and PSA measures appeared to have strong reliability and validity, as did the experimental measure for this study, SNF. However, the IPA appeared to have the weakest technical backing (Elliot et al., 2001). As with the previous Elliot study (1997), weaker correlations were found between the LNF and PSA measures with the criterion measures than those obtained by Kaminski and Good (1996). Like the previous Elliot (1997) study, the LNF and SNF tasks yielded the highest reliability and validity estimates. The authors concluded by noting that the study supported the use of DIBELS-M for identifying at-risk kindergartners for later reading difficulties, using the measures as a progress monitoring device, and for evaluating the effectiveness of pre-reading instruction (Elliot et al., 2001).

A study by Hintze, Ryan, and Stoner (2003) addressed the concurrent validity of the DIBELS as well as the diagnostic accuracy of the DIBELS in the prediction of performance

on the Comprehensive Test of Phonological Processing (CTOPP) among 86 students during their kindergarten year. The authors utilized three alternative form probes for each of the three DIBELS measures of LNF, ISF, and PSF tasks. The CTOPP measure included subtests such as Elision (requires the child to verbally state a word after the omission of a certain phoneme), Rapid Color Naming, Blending Words, Sound Matching, Rapid Object Naming, Memory for Digits, and Non-Word Repetition (Hintze et al., 2003). The authors found that the three DIBELS measures correlated strongly with the CTOPP subtests that purport to measure phonological awareness and had weak correlations with the CTOPP tasks that purport to measure rapid naming ability (Hintze et al., 2003). Additionally, the ISF and PSF tasks appeared to have the strongest correlations with the CTOPP tasks of Elision, Sound Matching, and Blending Words.

In a study conducted by Craycroft (2001), the test-retest reliability of the DIBELS PSF, ISF, and LNF tasks were evaluated. The study consisted of 152 kindergarten students with an age range of 5 years, 3 months to 7 years, 0 months. The inter-rater reliabilities for the DIBELS measures ranged from 97% for the ISF task, 69% for the PSF task, and 100% for the LNF task. The DIBELS PSF task was the only task that failed to meet the .80 criteria for an acceptable inter-rater reliability coefficient. In addition, the author found that the raw scores for the ISF task remained fairly stable from one test administration to the next, while the PSF and LNF tasks had significant point differences from one test administration to the next. The author noted that the scoring difference between the two test administrations for the PSF task may be attributable to the difficulty of scoring whether or not a student said each individual phoneme. The author goes on to suggest that additional training for scoring

the task may be prudent. The test-retest reliability coefficients for the tasks were PSF .82, ISF .85, and LNF .88, indicating a high level of test-retest reliability.

Midden (2003) conducted a follow-up study of Craycroft's (2001) participants. Of the original 152 kindergartners, reading achievement scores of 104 students were obtained in second grade. The reading achievement scores of the participants were determined by the California Achievement Test-Fifth Edition (CAT-5). The author found that the DIBELS measures significantly correlated with later reading ability with a coefficient of .77. Additionally, Midden (2003) found that the DIBELS LNF task resulted in a higher correlation with reading than the PSF task and ISF task, although the differences between the correlations were not significant.

In sum, the DIBELS measures appear to have adequate criterion related validity and inter-rater reliability. However, estimates of concurrent validity vary from one study to another. Elliot (1997) noted lower concurrent validity estimates for the LNF and PSF tasks than those found by Kaminski and Good (1996). Additionally, Elliot found the LNF and experimental SNF tasks to be the strongest measures, while Kaminski and Good stated that the LNF task appeared to be the strongest measure, while the PSF task was the most sensitive to change. Midden (2003) also found the LNF task to have the highest correlation with later reading ability. Craycroft (2001) found that the PSF task was the only measure among three DIBELS measures failing to meet the criterion for significant inter-rater reliability. However, Craycroft (2001) found high levels of test-retest reliability for all three DIBELS measures.

DIBELS as a Classification Tool

DIBELS measures allow educators and school psychologists to determine which children require additional skills and training to become successful readers. DIBELS is typically utilized through screening children in kindergarten and first grade at three key points in the year, fall, winter, and spring (Kaminski & Good, 1996). The measures allow for children who are not on target to becoming successful readers to be identified at multiple points in the academic year. Thus, the DIBELS can be used to identify which students are at-risk for later reading difficulties.

Good, Simmons, and Kame'enui (2001) used a set process for attaining benchmark goals for each of the DIBELS tasks. Benchmarks act as guidelines as to where a student's minimum ability should be during a given point in the school year. The benchmarks provide a guideline for identifying those students who are in need of intensive intervention, those who need some type of intervention, and those who are on the path to becoming successful readers. The benchmarks also allow for a comparison of where the student currently is to where they should be, providing a goal for the student to attain. As an example, the PSF task was administered to 78 students in the spring of their kindergarten year. One year later, in the spring of the first grade year, the students were assessed on a curriculum-based measure of oral reading fluency. A minimum score of 40 words read correctly per minute was judged as a desirable level of reading ability for a first grade reader. Conversely, a score at or below 10 words per minute indicated that a student lacked the appropriate skills to be a successful reader.

The authors' next step in determining appropriate benchmarks for the PSF DIBELS measure was to examine the PSF scores of students who performed successfully (reading 40

or more words) and unsuccessfully (reading 10 or fewer words) on the CBM measure (Good, Simmons, Kame'enui, 2001). Three groups were established based on the PSF scores. The first group of students scored 35 or better on the PSF task. Of those students, all but one was also successful on the CBM measurement. The second group was in the middle range with scores between 10 and 35 on the PSF task. The authors were unable to make clear predictions of this group's reading ability; a portion of the group proved successful on the reading measurement while others in the group did not. The third group of students achieved scores of 10 or less on the PSF measure. Of the group, only two out of 18 students were successful on the CBM reading task. From these data, the authors selected a PSF score of 35 or above as an indicator that the child will be successful at future reading attempts. Additionally, the students with scores at or below 10 on the PSF task were deemed as being in need of intensive reading intervention.

From their study, Good Simmons, and Kame'enui (2001) categorized each group of children's performance into to four different "zones" of ability or projected track of achievement. The first zone, termed "Zone A," refers to children who attained the earlier benchmark goal and later achieving the follow up benchmark goal (see Figure 1). These children are considered to be on the path to being successful readers. The remaining three zones represent children who have struggled at one or both benchmark goals. The students within "Zone B" were able to meet the first benchmark goal but failed to achieve the following goal. These children are seen as previously being on target but have since failed to progress with the appropriate skills to be a successful reader. The third zone, "C," contains students who did not attain the initial benchmark but later attained the follow-up benchmark. The students were previously at-risk of not reaching their reading goal but have become

successful readers. The final zone, “D,” encompasses those students who did not attain the initial benchmark goal and did not attain the later reading benchmark. In sum, the authors concluded that the performance of students across benchmark goals were accurate and consistent enough to utilize the DIBELS measures to later predict reading performance in relation to the fluency of skills in the acquisition of reading based on test administration at three points across the school year.

An example of the “zone” approach was presented in the study conducted by Hintze et al. (2003). The study measured the sensitivity of the DIBELS measures to identify those students who would fall into “Zone D.” These students were identified as having difficulty in reading based on the DIBELS measure. Specificity was also included in the Hintze et al. (2003) study, which refers to the ability of the DIBELS to identify students who do not have a reading problem. Each student is identified based on his or her achievement of the proper DIBELS benchmark for that grade and that point in the school year. According to Good Simmons, and Kame’enui (2001), these students would fall into “Zone A”—those that achieved consecutive benchmarks. False negatives would fall into “Zone B.” False negatives are students who pass an earlier benchmark but then fail a later benchmark. False positives are students in “Zone C” who were identified as having reading difficulty based on the failure of passing a DIBELS benchmark but passed subsequent benchmarks.

Hintze et al. (2003) concluded that both the DIBELS ISF and PSF measures correlated most strongly to the phonemic awareness composite of the CTOPP, as well as the phonological memory composite. However, specificity of the ISF task, as well as the PSF task, appeared low, which suggest that the tasks will over identify students as having deficits in phonological awareness skills, when in actuality, they may not. Additionally, the ISF and

PSF measures appeared to poorly predict which students were likely to have problems on the CTOPP measure; however, both measures were again highly capable of predicting which students would *not* have difficulty of the CTOPP measures. The authors concluded that the ability of the ISF and PSF measures to correctly classify a student based on the DIBELS cut-offs set by the manual are no better than chance. Hintze et al. (2003) suggested that the varied patterns of correlation with the CTOPP may be due to the fact that the children in the study had yet to develop the ability to consistently segment phonemes.

Hintze et al. (2003) noted that only one-third to one-half of the students may be correctly identified as either exhibiting difficulties in phonological awareness or being on target in the development of phonological skills. An implication of this finding is that while the specified DIBELS cut-offs are proficient in identifying students who demonstrate difficulty with phonological awareness skills, students are also over-identified as having difficulties in phonological awareness. However, the authors suggest that it is more prudent to over-identify students as having reading difficulties and provide extra instruction than it is to miss the identification of students who require more instruction but are not properly identified (Hintze et al., 2003).

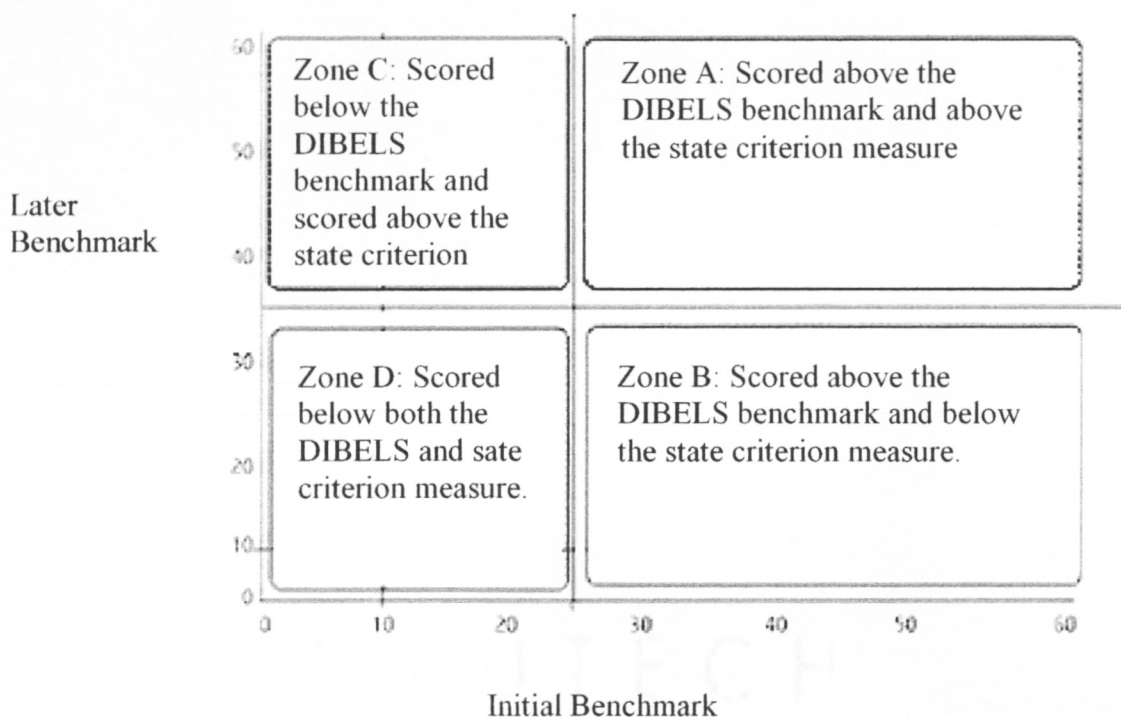


Figure 1: Classification of benchmark zones.

Predicting Performance on State Measures

In a study conducted by McGlinchey and Hixson (2004), 1,362 students were assessed over an 8-year period on both a curriculum-based measure of oral reading fluency (ORF) as well as the Michigan Educational Assessment Program's (MEAP) fourth-grade reading assessment. The purpose of the study was to examine the predictive validity of a CBM reading probe to the MEAP reading assessment. The CBM consisted of a 1-minute reading probe for the first 5 years of assessment and three, 1-minute reading probes for the last three years. All students were administered the same reading passages 2 weeks prior to the administration of the MEAP. A satisfactory score on the MEAP was a scaled score of 300 or above on each section. Students scoring 299 or below on both sections were deemed in the "Low" range. A score of 100 words correct per minute (WCPM) was chosen as a cut-off score for the oral reading fluency rate based on previous research. The authors examined

sensitivity, specificity, positive and negative predictive power, as well as the overall correct classification of the CBM measure.

Criterion-related validity was also examined with a fairly consistent relationship found between reading rate and MEAP scores, ranging from .49 to .81. Additionally, with a cut-off score of 100 words per minute correct, the specificity (students who scored above 300 on the MEAP as well as met the cut-off of 100 WCPM) was 74%. Positive predictive power (the correct identification of students who achieved below satisfactory scores) was 77%. Negative predictive power (the correct identification of students who achieved satisfactory scores) was 72%. Each measure was found to be above the base rate of 46% of achieving a satisfactory score. The overall correct classification was 74%. McGlichey and Hixson (2004) noted that school districts can raise or lower cut-off scores of CBMs or similar measures in order to adjust for the perceived level of ability within the school district. When using a higher cut-off score the probability of predicting a satisfactory score (versus a mastery score) will likely increase, but the likelihood of predicting a failure will also decrease.

Recent studies have begun to utilize the DIBELS measures in an attempt to assess reading ability specific to that area or school district. A study conducted by Linner (2003) in the Alaska district of Kenai Peninsula Borough incorporated DIBELS and a CBM task to predict reading ability in third grade. Linner collected data over the spring of four years. He then analyzed the data of each CBM measure for each school year with the students' benchmark performance. In general, Linner found that the measures had high negative predictive power (98%)—meaning that if a student was deemed not to have a problem in reading based on a given DIBELS benchmark performance, he or she was also classified as not having a problem on the CBM measure (Linner, 2003). Conversely, the positive

predictive power of the measures was not as strong (24%), indicating that there was some disagreement as to which students did in fact have a problem in reading. Specifically, the author found that the DIBELS LNF task correctly categorized students overall about half the time based on their third grade ORF scores.

Purpose

Reading is a critical skill in the repertoire of a student; without the skill a student is likely to fall behind his or her peers. Early identification of those students who struggle with reading is even more important when considering the short, critical period of time when reading develops. As Juel (1988) noted, a child who is a poor reader at the first grade has a .88 chance of remaining a poor reader at fourth grade. Thus, it is essential that students who struggle with reading be identified early; however, such identification is beneficial only if the measures used are valid and reliable. One possible measure which can be employed for the early identification of students with reading problems is the DIBELS measure.

An examination of the literature on the DIBELS measures indicated that, for the most part, the tasks are succinct measures that are valuable for the identification of students who are “at-risk” for failing to develop appropriate reading skills (Elliot, 1997; Elliot et al., 2001; Hintze et al., 2003; Kaminski & Good 1996; Midden, 2003). The ISF, LNF, and PSF measures have been shown to accurately predict later reading skills; however, some researchers (Craycroft, 2001; Elliot, 1997) noted that the PSF task may not be as strong a measure due to the difficulty in scoring each phoneme as the child segments a word. Thus, while the DIBELS measures appear to be successful in the prediction of later reading skills, more information is needed to determine the accuracy with which the measures classify students as either “at-risk” or “on-track” for the development of reading skills.

Good and Kaminski (2002) provide specific cut-off scores for the DIBELS measures to indicate whether or not a student is “at-risk” for later reading difficulties. The first purpose of this study was to evaluate the effectiveness of the DIBELS cut-off scores for identifying local students “at-risk” for scoring low on state-mandated tests of reading. Craycroft (2001) assessed a sample of 152 kindergarten students in a school district in a southern state with the DIBELS. Midden (2003) conducted a follow-up study when the same students were in the 2nd grade to determine the correlation between the DIBELS and the CAT-5. This research examined the previously collected data to evaluate the published cut-off scores.

The second purpose of this study was to set specific cut-off scores on the DIBELS for the school district in a southern state. The 10th and 25th percentiles were identified on the CAT-5 reading achievement test. Cut-off scores on the DIBELS were then set based on the two levels of performance on the CAT-5.

Method

Participants

The participants were the follow-up sample of 104 students from a school district in a southern state obtained by Midden (2003). The sample included 57 (54.8%) males and 47 (45.2%) females when they were in 2nd grade. Most of the students were Caucasian (74.1%) with 18.3% African-American, 1.9% Hispanic, 3.8% Asian, and 1.9% classified as other. Although Midden (2003) had 48 fewer participants than the original kindergarten sample obtained by Craycroft (2001), Midden assumed her sample was representative of the original sample based on similar demographics of each group. Permission to conduct the original study was obtained by Craycroft (2001) from the school district's board of education and Western Kentucky University's Human Subjects Review Board.

Materials

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) is a set of standardized, individually administered measures designed to assess early literacy skills. The number of tasks given is dependent on the child's grade in school and the point in the school year in which the child is being tested. Beginning in the fall of the kindergarten year, the Letter Naming Fluency (LNF) task of the DIBELS is administered. The measure requires each student to name as many random upper and lower case letters as possible within one minute. The student's score is the number of correctly named letters.

The Phonemic Segmentation Fluency (PSF) task and Initial Sound Fluency (ISF) are designed to assess phonemic awareness. The ISF task (formerly known as Onset Recognition

Fluency) assesses a child's ability to differentiate the initial sound of a given word from the rest of the word (Kaminski & Good, 1998). The child is presented with a page consisting of four pictures. The examiner then states the names of the pictures and asks the child to point to or state the picture that begins with a certain sound orally presented by the examiner. The examiner calculates the amount of time taken to identify/produce the correct sounds and converts the score into the number of initial sounds correct in a minute. However, Craycroft (2001) obtained only the raw scores on the ISF task (vs. rate per minute) due to poor inter-rater scoring agreement.

The Phoneme Segmentation Fluency (PSF) task assesses the student's ability to segment or break down three- and four-phoneme words into the individual phoneme components. The examiner orally presents words one at a time and then requires the child to break the word down into individual phonemes and recite the phonemes to the examiner (Good & Kaminski, 2000). A student's score is determined by the number of correct segmentations provided in one minute (Good & Kaminski, 2002).

The California Achievement Test-Fifth Edition (CAT-5; CTB McGraw-Hill, 2002) is a norm-referenced, group-administered achievement test. The test was developed by Macmillan/McGraw-Hill and published by CTB/McGraw-Hill Assessment Products and Services. The test is designed to measure achievement in the basic achievement skills taught in school including reading, spelling, mathematics, study skills, science, and social studies for students in kindergarten through 12th grade (CTB McGraw-Hill, 2002). At the second-grade level the CAT-5 includes three reading subtests: (a) Word Analysis, (b) Vocabulary, and (c) Comprehension. For the purpose of this study the Reading Total was used.

Procedure

One hundred and fifty-two students were initially assessed with the DIBELS during January of their kindergarten year (Craycroft, 2001). The present study examined the CAT-5 reading achievement scores for 104 students during the Fall semester of their second grade year obtained by Midden (2003) from school district personnel. The present study further analyzed the data collected by Midden (2003) and Craycroft (2001).

The first goal of the current research was to examine the published DIBELS benchmark on the second grade CAT-5 reading test results. Data were analyzed by initially using scatter plots to compare DIBELS scores on the LNF and PSF tasks to the CAT-5 Reading Total scores of the 104 participants. For the first goal of the study, the ISF and DIBELS Total were not analyzed. Craycroft (2001) obtained only raw data on the ISF subtest rather than converted scores, and a benchmark is not available for raw data. The developers of the DIBELS measures do not total the scores from individual measures, thus a benchmark score for the DIBELS Total is not available. DIBELS scores were plotted on the x-axis and corresponding CAT-5 scores were plotted on the y-axis. Two cut-off scores were then applied to the CAT-5 Reading Total scores for each DIBELS test comparison. A more lenient cut-off at the 25th percentile was selected to identify those students scoring in the lower quartile on the CAT-5 Reading Test. A score of 586 is at the 25th percentile according to the CAT-5 Fall Norms Book (CTB McGraw-Hill, 1992). Past studies by Snider (1997) and Hintze et al. (2003) reported that the DIBELS measures over-identified students as being at-risk for reading difficulties. For this reason, a second, more conservative, cut-off score at the 10th percentile was also applied when examining the data. For the CAT-5 Reading Total, a score of 558 is at the 10th percentile. These procedures resulted in four scatter plots due to

two DIBELS subtests (LNF and PSF) and two levels of cut-off scores (10th and 25th percentiles) on the CAT-5 . However, after examining the scatter plots for both the LNF and PSF tasks at the 10th and 25th percentiles, it was determined that the change in criterion (from a more conservative to a more lenient cut-off) resulted in very little difference in the categorization of students. Thus, a compromise criterion at the 16th percentile for the Reading Total (570) was selected as the criterion for all comparisons. The 16th percentile was chosen because it is one standard deviation below the mean and falls between the initial criteria at the 10th and 25th percentiles.

After each of the CAT-5 cut-off scores were applied to the four scatter plots, the data were examined based on the provided DIBELS benchmarks. Benchmarks act as guidelines as to where a student's minimum ability should be during a given point in the school year (Good & Kaminski, 2002). The percent of false negatives was determined using the published DIBELS benchmarks. New cut-off scores were then determined based on a criterion of less than 5 percent false negatives, and the rate of false positives and false negatives was re-examined.

Results

The published at-risk benchmark for the LNF task is less than 15 letters named correctly in one minute, and for the PSF task a score of less than 7 correct phonemes in one minute is considered at-risk (Good & Kaminski, 2002). The rate of false negatives and of false positives were then determined based on the provided DIBELS benchmarks. False negatives are students who scored at or above the DIBELS benchmark but fall below the CAT-5 cut-off criterion at the 16th percentile. Thus, of the students who scored below the 16th percentile, false negatives are those students who scored above the at-risk benchmark. False positives are those students who scored less than the DIBELS benchmark but scored above the 16th percentile on the CAT-5 Reading test.

For the LNF task, a comparison at the 16th percentile on the CAT-5 Reading Test and the DIBELS cut-off of less than 15 letters named correctly resulted in a false negative rate of 60% (of the students who scored below the 16th percentile on the CAT-5 Reading Total, 60 percent were false negatives) with a false positive rate of 9% (of the students who scored above the 16th percentile on the CAT-5 Reading Total, 9 percent were false positives). See Figure 2. The comparison of the PSF task at less than 7 correct segmentations in a minute and a criterion score on the CAT-5 at the 16th percentile resulted in 20% false negatives and 28% false positives (see Figure 3).

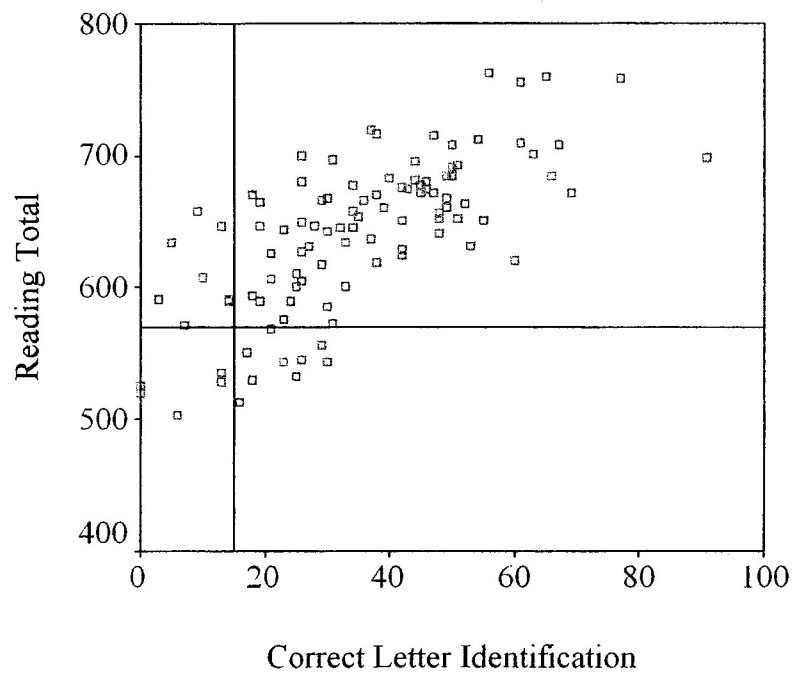


Figure 2. Classification of students based on the published LNF benchmark of 15 correct letters named in one minute.

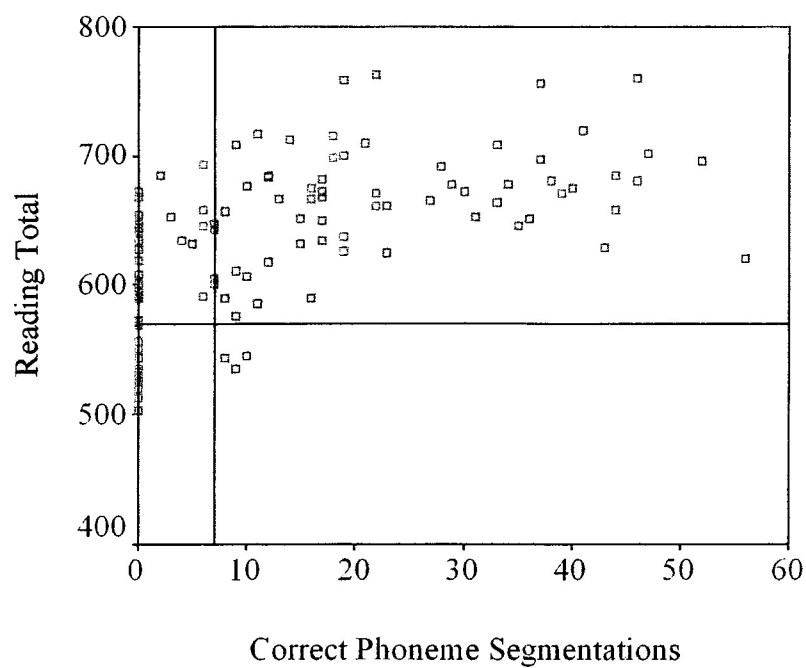


Figure 3. Classification of students based on the published PSF cut-off of 7 correct phoneme segmentations in one minute.

To attain the second goal of the study, the benchmarks for each DIBELS measure was then adjusted to a criterion which identified less than 5 percent of the students as false negatives. The criterion was chosen to restrict the number of students identified as being on-target by the DIBELS, but who still fell below the 16th percentile on the CAT-5 Reading test.

For the LNF task, the published DIBELS cut-off was adjusted to a criterion of 31 or fewer letters named correctly in one minute (see Figure 4). Once again the comparison was made with the CAT-5 criterion at the 16th percentile. Following the adjustment, the rate of false negatives fell from 60% with a cut-off at 15 correct letters named in a minute to 0% false negatives. Additionally, the rate of false positives rose from 9% to 36% with the adjusted cut-off score.

The PSF task was also analyzed with a criterion of less than 5% false negatives. The published DIBELS cut-off score was adjusted from 7 correct segmentations in a minute to 11 correct segmentations in a minute (see Figure 5). The adjustment resulted in a decrease in false negatives from 20% with the published benchmark to 0% false negatives following the adjustment. The rate of false positives increased from 28% to 40% with the criterion set at 11 correct segmentations in a minute.

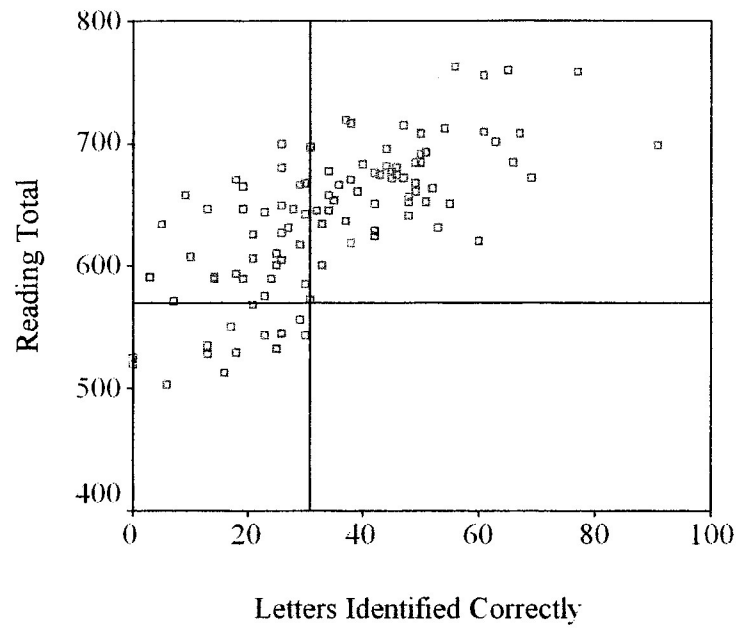


Figure 4. Classification of students based on the adjusted LNF cut-off at 31 correct letters named in one minute.

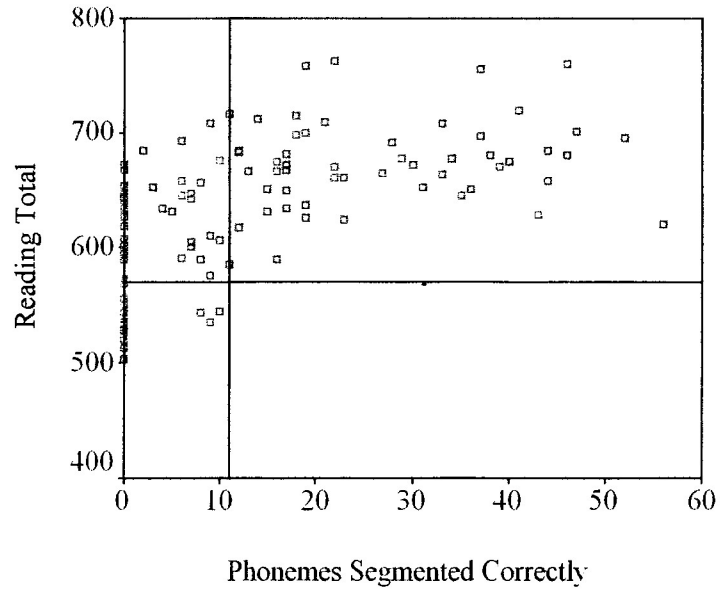


Figure 5. Classification of students based on adjusted PSF cut-off at 11 correct phoneme segmentations in one minute.

ISF scores were reported in raw data scores rather than converted rate scores during the initial data collection, thus a DIBELS published benchmark was not available. To establish a false negative rate of less than 5%, the cut-off criterion was set at 17 words read correctly. This criterion resulted in a false negative rate of 0% and a false positive rate of 100%, based on students who scored above the 16th percentile (see Figure 6).

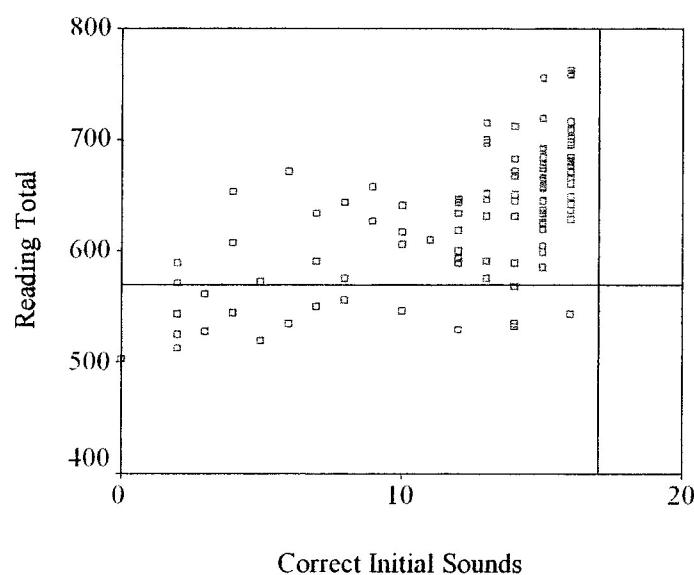


Figure 6. Classification of students using a cut-off of 17 correct initial sounds based on less than 5% false negatives.

As noted, the DIBELS does not currently utilize a DIBELS Total score, thus the initial comparison was made using the CAT-5 criterion at the 16th percentile and a DIBELS cut-off which identified less than 5% false negatives. The criterion was determined to fall at a score of 55 when students' scores on the LNF, PSF, and ISF tasks are summed. The cut-off score resulted in 0% false negatives, with a false positive rate of 33% (see Figure 7).

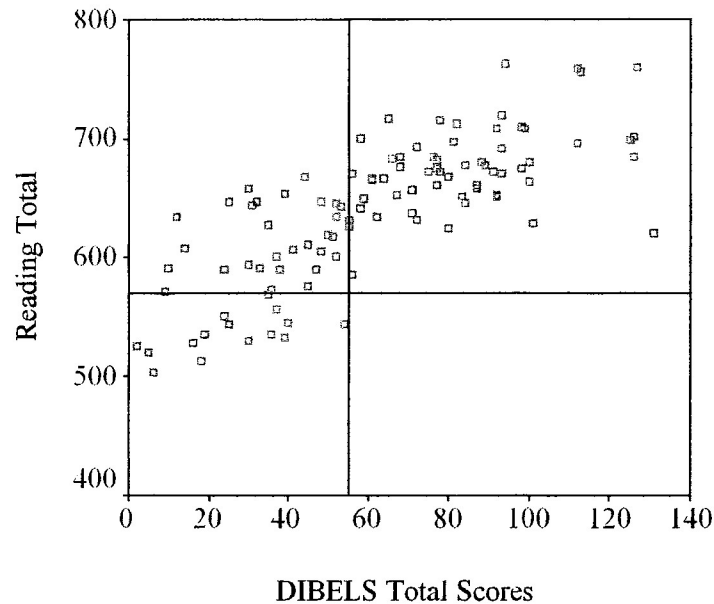


Figure 7. Classification of students based on DIBELS Total cut-off of 55.

Table 1 presents the published and adjusted false positives and false negative rates for all comparisons, as well as additional statistical comparisons. An interesting point to examine is the effect of the benchmark adjustment to the Overall Correct Classification (OCC) of each task. Although the benchmark adjustments resulted in a decrease in false negatives, the OCC also decreased for each task. The OCC for LNF decreased from 84% to 69%. Similar results were also noted for the PSF task. The OCC for the PSF task decreased from 73% to 65%. Another important note in the examination of the OCC are the results for the DIBELS Total. With an OCC of 72%, the DIBELS Total appears to be an adequate measure in the classification of students as at-risk.

Table 1

Predictive Accuracy of Published and Adjusted DIBELS Cut-off Scores

T	LNF		PSF		ISF	DIBELS
	PUB	ADJ	PUB	ADJ	ADJ	ADJ
FN	.60	.00	.20	.00	.00	.00
FP	.09	.36	.28	.40	1.00	.33
SENS	.40	1.00	.80	1.00	1.00	1.00
SPEC	.91	.64	.72	.60	.00	.67
OCC	.84	.69	.73	.65	.16	.72
PPP	.43	.32	.32	.29	.15	.34
NPP	.90	1.00	.96	1.00	.00	1.00

Note. The DIBELS does not provide published cut-off scores for ISF raw data or for a DIBELS Total. T = total; PUB = published; ADJ = adjusted; FN = false negatives; FP = false positives; SENS = sensitivity; SPEC = specificity; OCC= overall correct classification; PPP = positive predictive power; NPP = negative predictive power.

As of yet, it appears that no research has looked at the effectiveness of the DIBELS in the reading classification of minority students. A post-hoc examination of the classification of the minority students in the current sample was conducted using the DIBELS Total score. The DIBELS Total was plotted with a cut-off score of 55 and the CAT-5 Reading Total cut-off at the 16th percentile (see Figure 7). Two different scatter plots were examined. The first looked at the overall sample *without* minority students, and a second scatter plot was created

using *only* the minority students. Statistical comparisons were then examined to determine the effect of the minority students on the classification of students.

Using the DIBELS cut-off at 55 without minority students resulted in a false negative rate of 0%. In addition, the false positive rate was very similar for both samples with a rate of 33% for the sample with the minority students and a rate of 30% for the sample without the minority students. The scatter plot in which *only* the minority students were examined resulted in rates similar to the sample with only the Caucasian students. The rate of false negatives remained at 0 percent; however, the rate of false positives was slightly higher with a rate of 44 percent. Table 2 presents a comparison of each of the sample configurations. As shown, the statistical comparisons are relatively consistent across groups. The rate of Positive Predictive Power (PPP) was slightly variable, indicating that the sample with the minority students only served as a better classification tool than the total sample and Caucasian sample.

Table 2

Comparison of DIBELS Total With and Without Minority Students

	Total Sample	Caucasian	Minority
FN	.00	.00	.00
FP	.33	.30	.44
SENS	1.00	1.00	1.00
SPEC	.67	.70	.56
OCC	.72	.71	.74
PPP	.34	.15	.61
NPP	1.00	1.00	1.00

Note. The DIBELS does not provide published cut-off scores for ISF raw data or for a DIBELS Total. FN = false negatives; FP = false positives; SENS = sensitivity; SPEC = specificity; OCC= overall correct classification; PPP = positive predictive power; NPP = negative predictive power.

Discussion

Overall, the published DIBELS measures appear to be an adequate device for identifying students in a school district in a southern state as either at-risk or on-target in reading. However, the measures tended to over-identify kindergarten students (in terms of false negatives) as being on-target when in actuality they were not. Because of this discrepancy, adjustments were made to decrease the number of students who were at-risk for later reading difficulties. While the adjustments did eliminate any false negatives, the rate of false positives greatly increased in some instances.

The rate of false negatives for the published LNF benchmark was 60%, with a false positive rate of 9%. The indication is that the task over identified students as being on-target for reading success when in actuality they were not. To decrease the rate of false negatives to less than 5% the DIBELS cut-off was adjusted from 15 correct letter identifications to 31 correct letter identifications. This adjustment resulted in no false negatives, but the rate of false positives increased from 9% to 36%. These results were expected because as the rate of false negatives decreases, the rate of false positives increases. However, it is reasoned that it is more prudent for students to receive intervention when it may not be needed, than to risk not providing early intervention to a student who is at-risk but is not identified as such. However, depending on the intensity of the intervention, it may be difficult (due to time and money constraints) to provide intervention to all the students determined to be at-risk. It may be more feasible to set a cut-off which decreases the rate of false negatives but does not eliminate all students from this zone as did the predetermined 5% criterion in this study. The

DIBELS publishers suggest a cut-off of 27 correct letters named in a minute for identifying students who are at *some* risk. This cut-off may be more appropriate as an at-risk cut-off for the Bowling Green City School District versus the published cut-off of 15 or the adjusted cut-off of 31 correct letters named used in the current study.

The same results were seen with the adjustments for the PSF task. The rate of false negatives decreased from 20% to 0%, while the rate of false positives increased from 28% to 40%. However, the adjustment to account for less than 5% false negatives was not as dramatic for the PSF task as it was for the LNF task. The adjusted cut-off was set at 11 correct phoneme segmentations per minute versus the published 7 correct segmentations per minute. This cut-off appears to be an appropriate one as it decreased the amount of false negatives to a rate of 0%; however, the rate of false positives was once again increased (from 28% to 40%).

The use of scatterplots to evaluate the DIBELS' predictive validity provides an interpretation that differs from past studies which have found the LNF task to have the strongest correlation with later reading abilities (Elliot, 1997; Kaminski & Good, 1997; Midden, 2003). The PSF task resulted in a lower percentage of false negatives when adjusted; however, the rate of false positives was higher for the PSF task than for the LNF task. The LNF task appears to over-identify students as being at-risk when they are not. Only when the published DIBELS criterion is more than doubled do the false negatives decrease to a rate less than 5%.

It is also important to note that the DIBELS measures are primarily meant to be utilized as a progress monitoring tool. As such, it may not be critical that a specific cut-off score be determined. A child who is deemed at-risk should be provided

intervention. However, the intervention need not be long term and expensive for *all* children. By using the DIBELS as a progress monitoring tool, teachers and others will know which children are making progress and which are not. As children achieve the benchmarks, the extra intervention services can be dropped. For those children who do not make adequate progress, more intensive interventions can be implemented with little time being lost.

The current study also looked at the ability of the ISF task to correctly identify students as being at-risk. The results indicated that the use of the ISF raw data was insufficient in the identification of students as at-risk—or as an identification tool at all. Using the 5% cut-off resulted in a rate of 0% false negatives, but the rate of false positives was 100 percent. Thus, no students were identified as on-target, and all students who were misidentified were identified as false positives. Additionally, no students scored above both the ISF cut-off and the 16th percentile on the CAT-5. It is prudent to remember that the task is not meant to be analyzed using raw data; therefore, if converted data had been available it is possible that the task may be more successful as an identification tool.

The current study also set out to examine the DIBELS Total as a possible identification tool. The total was obtained by summing the students' scores on each of the DIBELS measures to garner one overall score. Once again, the cut-off was set at a point to identify less than 5% false negatives at the 16th percentile on the CAT-5. Using this cut-off, 0% false negatives were identified and 33% of false positives were identified. It appears that the DIBELS Total may be a useful measure in identifying students who are at-risk for reading difficulties. Of the students who scored below the

adjusted cut-off, all were correctly identified. Only 33% of students were identified as needing intervention when they actually ended up being on-target. The false positive results were only slightly higher than those of the published DIBELS benchmarks for the LNF and PSF task and were lower than the adjusted cut-offs for both tasks, indicating that the DIBELS Total was a stronger classification tool than the adjusted LNF and PSF tasks.

Finally, the DIBELS Total was also examined as a classification tool with minority students. It appears that the DIBELS Total is equally strong at identifying students as at-risk when minority students are included in the sample, as when they are not. These results indicated that minority students are classified equally as well as non-minority students. In fact, the DIBELS Total was slightly more successful as an identification tool for the minority students than for the total sample. Therefore, even though research has not been conducted on the use of DIBELS with minority students, it appears, at least in the case of the DIBELS Total, the task is a strong measure in the identification of students as being at-risk.

Limitations

Although the current study provided additional research evidence for the effectiveness of DIBELS as a classification tool, there are four limitations that may affect the results of the present study. The first limitation is that the conclusions made about the DIBELS in this study are based on the assumption that the CAT-5 is a reliable and valid measure of a child's reading ability. Other measures of reading achievement may be superior.

A second limitation of the study is that converted rate scores were not available for the ISF data, thereby preventing the data from being analyzed based on the published DIBELS benchmark and as an adjusted benchmark. Their absence also prevents an appropriate interpretation of the DIBELS Total score. Because the DIBELS total score determined in this study was based on the ISF raw data, it is unknown what the appropriate total cut-off score would be based on converted ISF data.

The third limitation involves the uncertainty of what phonemic awareness or general reading instruction was provided following the DIBELS administration and prior to the CAT-5 administration. All students presumably received reading instruction. Phonemic awareness strategies or specific reading instruction interventions may also have been provided to some or all of the students before the students were assessed on the CAT-5 in their second-grade year. Therefore, it is unclear whether students were actually misidentified by the DIBELS as being at-risk when they were not, or if the at-risk students received instruction and interventions that resulted in their overcoming actual reading deficits by the time of the CAT-5 administration.

The fourth limitation is the relatively small sample size and limited diversity which limits the generalizability of the results. The sample consisted of 104 students, almost evenly divided between males and females. However, the sample consisted of a mostly Caucasian population (74%). In addition, English Language Learner (ELL) students were included in the minority group. Thus, native English speakers were analyzed in the same group as students who may not have been fluent in English.

Inclusion of ELL students in the minority sample may affect the interpretation of the results.

Future Research

The DIBELS is a relatively new measure (first published in 1996) and as such there is a relatively small amount of research involving the measure. However, the usage of the DIBELS is growing and the research base is growing as well. The current study serves as a catalyst for future research in several areas. The first area of future research is the examination of the DIBELS Total score. As noted, the DIBELS Total appears to be a strong measure in the classification of students as at-risk in reading. The Total proved to be a better measure than any of the individual DIBELS subtests. Currently, no other studies have examined the usefulness of a DIBELS Total score. As such, more research would be prudent. Future research should be directed towards looking at the DIBELS Total with converted scores rather a combination of raw and converted data, as was the case with the present study. The DIBELS Total can also be examined with other grades as well. For instance, the tasks assessed at the first grade level include the LNF, PSF, ISF tasks, as well as the Nonsense Word Fluency (NWF) task. Scores from each of these tasks can be summed to establish a DIBELS Total; however, one would also have to determine an appropriate DIBELS benchmark for the first grade tasks based on a chosen criterion.

As noted, no research currently exists on the utility of the DIBELS with minority students. Some studies have been conducted with minority students and curriculum based measurement (CBM), which shares similar principles as the DIBELS; however, the research is somewhat mixed at this point. In an initial study

conducted by Baker and Good (1995), it was found that an English language CBM was just as reliable and valid for bilingual students as for English-only students. Another study found that CBM tended to over-identify African American students as having reading difficulties as compared with Caucasian students (Kranzler, Miller, & Jordan, 1999). However, a follow-up study found that the results of the Kranzler et al. research was not methodologically sound, and in fact, CBM neither over- or under-identified African American students (Hintze, Callahan, Matthews, Williams, & Tobin 2002). The present study found that the DIBELS Total appeared to be an effective tool in the classification of minority students as at-risk or on-target; however, further research is needed to truly understand the effectiveness of the measure using larger samples.

References

- Adams, M. J. (1990) *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Baker, S. K., & Good, R. H. (1995). Curriculum-Based Measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. *School Psychology Review, 24* (3), 537-560.
- Chafouleas, S. M., & Martens, B. K. (2002). Accuracy-based phonological awareness tasks: Are they reliable, efficient, and sensitive to growth? *School Psychology Quarterly, 17*, 128-147.
- Coyne-Martinson, P. A. (2000). Assessing the prereading language development of kindergarten students: *The reliability, validity, and sensitivity of Basic Language Assessment Story Tasks as dynamic indicators of basic early literacy skills*. (Doctoral dissertation, University of Oregon, 2000). *Dissertation Abstracts International, 60*, 3603.
- Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of test bias in curriculum-based measurement of reading across racial/ethnic groups. *School Psychology Quarterly, 14*, 327-342.
- CTB McGraw-Hill (1992). *California Achievement Test* (5th ed.) Fall Norms Book: September through November. Monterey, CA.
- CTB McGraw-Hill (2002). *California Achievement Test* (5th ed.). Monterey, CA.
- Craycroft, L. A. (2001). *Test-retest reliability of phonemic awareness assessment instruments on kindergarten students*. Unpublished Specialist project, Western Kentucky University, Bowling Green, KY.

- Department of Education. (2004). *No Child Left Behind provision gives schools new flexibility and ensures accountability for students with disabilities*. Retrieved May 14, 2004, from <http://www.ed.gov/nclb/freedom/local/specedfactsheet.pdf>
- Digest of Education Statistics. (2002). *Children 3 to 21 years old served in federally supported programs for the disabled*. (1976-77 to 2000-01). Retrieved July 18, 2004, from <http://nces.ed.gov/programs/digest/d02/tables/dt052.asp>
- Elliot, J., Lee, W. L., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills—Modified. *School Psychology Review*, 30, 33-49.
- Elliot, J. K. (1997). A reliability and validity study of DIBELS. (Doctoral dissertation, University of Kansas, 1997). *Dissertation Abstracts International*, 435.
- Fleming, K. M. (2000). The effect of instruction, rapid automatized naming, intellectual phonological awareness skill, and age on phonological awareness growth trajectories. (Doctoral dissertation, University of Oregon, 2000). *Dissertation Abstracts International*, 60, 3604.
- Good, R. H., & Kaminski, R. A. (1996). Assessment of instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly* 11, 326-336.
- Good, R. H., & Kaminski, R. A. (Eds.) (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th edition). Eugene, OR: Institute for the Development of Educational Achievement.

- Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5, 257-288.
- Griffith, P. L., & Olson, M. W. (1992). Phonemic awareness helps beginning readers break the code. *The Reading Teacher*, 45, 516-523.
- Hempenstall, K. (1997). The role of phonemic awareness in beginning reading: A review. *Behaviour Change*, 14, 201-214.
- Hintze, J. M., Callahan, J. E., Matthews, W. J. Williams, S. A. S., & Tobin, K. G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review*, 31, 540-553.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review*, 32, 541-556.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology* 80(4), 437-447.
- Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem-solving model: Dynamic Indicators of Basic Early Literacy Skills. In M. R. Shinn (Ed.) *Advanced applications of curriculum-based measurement* (pp. 113-142). New York: Guilford Press.

- Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*, 215-227.
- Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). *The alphabetic principle and learning to read. Phonology and reading ability*. Ann Arbor, MI: The University of Michigan Press.
- Linner, S. (2003, April). DIBELS and CBM as a predictor for the Alaska high stakes assessment third grade reading. In R. H. Good (Chair), *DIBELS and reading CBMs used as a predictor for high-stakes achievement tests*. Symposium conducted at the meeting of the National Association of School Psychologists, Dallas, TX.
- Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.
- MacDonald, W. G., & Cornwall, A. (1995). The relationship between phonological awareness and reading and spelling achievement eleven years later. *Journal of Learning Disabilities, 28*, 523-527.
- Mann, V. A. (1993). Phoneme awareness and future reading ability. *Journal of Learning Disabilities, 26*, 259-269.
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33*, 193-203.
- Mercer, C. D., & Mercer, A. R. (2001). *Teaching students with learning problems* (6th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

- Midden, S. P. (2003). *Predictive validity of kindergarten awareness measures on second grade reading skills*. Unpublished Specialist project, Western Kentucky University, Bowling Green, KY.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved February 21, 2004, from <http://www.nichd.nih.gov/publications/nrp/smallbook.htm>.
- O'Connor, R. E., Jenkins, J. R., Leicester, N., & Slocum, T. A. (1993). Teaching phonological awareness to young children with learning disabilities. *Exceptional Children*, 59, 532-546.
- Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology*, 76, 1309-1324.
- Shinn, M. R. & Bamonto, S. (1998). Advanced application of Curriculum-Based Measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of Curriculum-Based Measurement* (pp.1-31) New York: Guilford.
- Snider, V. E. (1995). A primer on phonemic awareness: What is it, why it's important, and how to teach it. *School Psychology Review* 24, 443-456.
- Snider, V. E. (1997). The relationship between phonemic awareness and later reading achievement. *Journal of Educational Research*, 90, 203.
- Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, 86, 221-234.

- Stanovich, K. E. (1986). Matthew effect in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 36-407.
- Torgesen, J.K., Alexander, A. W., Wagner, R.K., Rashotte, C.A., Voeller, K., Conway, T. & Rose, E. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional programs. *Journal of Learning Disabilities*, 34, 33-58.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A. (1994) Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities* 27, 276-286.
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvin, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, 91, 579-593.
- Torgesen, J. K., Rashotte, C. A., & Alexander, A. (2001). Principles of fluency instruction in reading: Relationships with established empirical outcomes. In M. Wolf (Ed.), *Dyslexia, Fluency, and the Brain* (pp. 333-355). Parkton, MD: York Press.
- Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology*, 40, 7-26.
- Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology* 77, 417-427.

- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101, 192-212.
- Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading and Research Quarterly*, 23, 159-177.
- Yopp, H. K. (1992). Developing phonemic awareness in young children. *The Reading Teacher*, 45, 696-703.
- Yopp, H. K. (1995). A test for assessing phonemic awareness in young children. *The Reading Teacher*, 49, 20-29.